

09/724857

FIGURE 1

SCHEMATIC OF A GENE

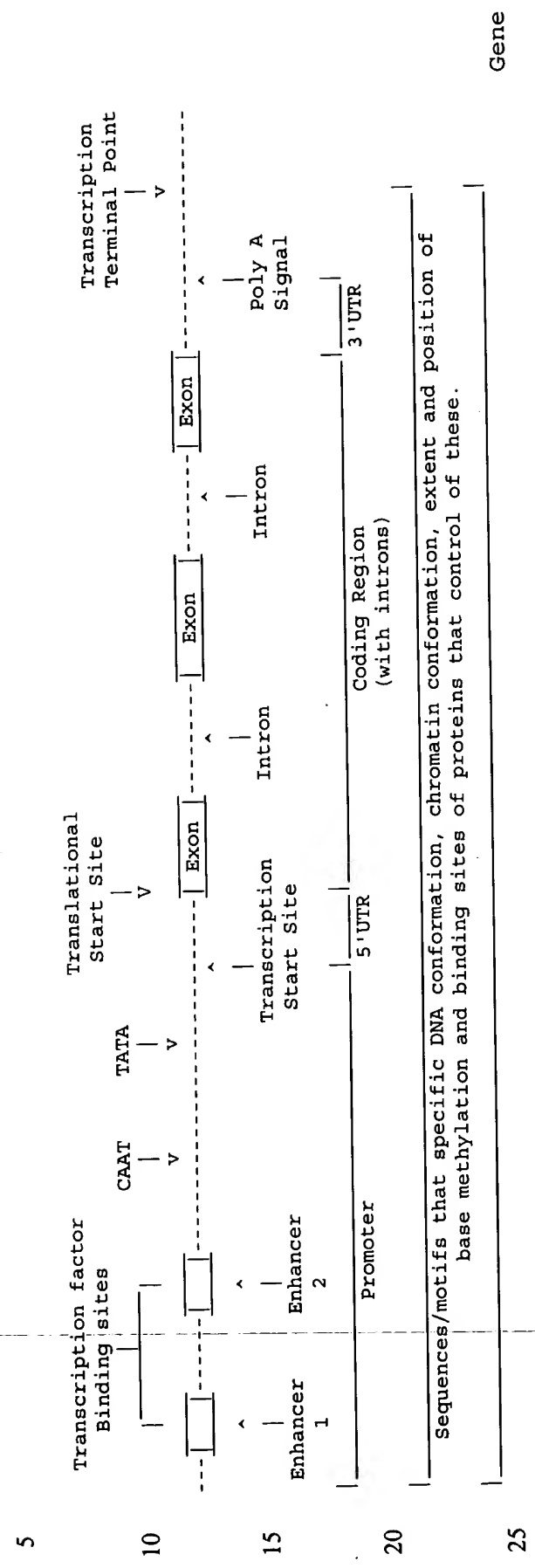


FIGURE 2

	-4242	GCATGCACTG	CCACAAGTAG	TGAACTCATG	GTTTTACCTC	CTCAAGTAGA
	-4192	AAACCTTTTG	AGTGAATTTG	AAGATTTATT	CTCCCAAGAA	GGACCCATTG
	-4142	GGCTTCCTCC	TCTTAGGGGG	ATAGAACATC	AAATTGACTT	TATACCGGGG
5	-4092	GCAAGCCTAC	CAAATAGGCC	TCCTTATAGA	ACCAACCCCG	AGGAAACAAA
	-4042	GGAGATAGAA	TCACAAGTTC	AAGACTTGTT	GGAGAAGGGT	TGGGTTCAAA
	-3992	AGAGCCTAAG	CCCTTGTGCT	GTACCTGTCT	TGTTGGTGCC	AAAAAAAGAT
	-3942	GGAAAAATGGC	GTATGTGTTG	TGATTGTAGA	GCAATCAACA	ACATCACCAT
	-3892	CAAGTATAGG	CATCCAATCC	CAAGGCTTGA	CGATATGCTT	GATGAATTGC
10	-3842	ATGGGTCAAC	TCTATTCTCC	AAAATTGACC	TTAAAAGTGG	ATATCACCAA
	-3792	ATTCGAATCA	AGGAGGGTGA	TGAGTGGAAA	ACCGCTTTTA	AGACCAAATT
	-3742	TGGATTATAT	GAGTGGTTGG	TGATGCCCTT	TGGTCTTACT	AACGCTCCAA
	-3692	GTACATTCAT	GAGGCTTATG	AATCACACCT	TGAGGGATTG	TATAGGTAAG
	-3642	TATGTAGTAG	TTTATTTTGA	TGATATCTTA	GTATATAGTA	AAACCTTAGA
15	-3592	AGACCATCTA	AGTCACCTTA	GGGAAGTTCT	TCTAGTTCTT	AGGAAAAATA
	-3542	GTCTTTTTCG	CAATAGGGAT	AAGTGTACCT	TTTGTGTAGA	TAGCGTAGTC
	-3492	TTTTTAGGCT	TTATAGTAAA	CCAAAAGGGG	GTGCATGTAG	ATCCCAGAGAA
	-3442	AATCAAAGCC	ATCCGCGAGT	GGCCAACTCC	ACAAAATGTA	AGTGATGTGA
	-3392	GAAGTTTTC	TGGGTTAGCT	AGCTTCTATA	GAAGGTTTGT	TCCCAATTTT
20	-3342	TCTAGCCTAG	CTTCTCCCTT	GAATGAACTT	GTAAAAAAG	ATGTTGCATT
	-3292	TTGTTGGAAT	GAAAAGCATG	AGCAAGCCTT	TCAAAGGCTA	AAAGCTCACT
	-3242	CACCAATGCA	CCCATCTTAT	CTCTTCCAAA	TTTTTCCAAA	CTTTTGGAGA
	-3192	TAGAGTGTGA	TGCATCGGGA	GTAGGCATAG	TGCGGTTTTC	TTGCAAGGTG
	-3142	GACACCCCTT	GCTTATTTTA	GTGAAAAACT	CCATGGTGCC	ACCCTCACTA
25	-3092	CCCCACCTAT	GACAAAGACT	CTATGCTCTT	GTGCGACCCT	AAAGACTTGG
	-3042	GGAACACTAC	CTTGnGTCCC	AAAGAAATTTG	GnTATCCATA	GTGATCACGA
	-2992	GTCTTTAAAA	TATTTAAAGG	GCCAACACAA	GCTCAATAAG	AGACATGCTA
	-2942	AATGGATGGA	ATTTCTTGAA	CAATTTCTCTT	ATGTCATCAA	ATACAAGAAA
	-2892	GGGAGCACCA	ATATAGTGGC	CGATGCTCTT	TCTAGACGGC	ACACTCTCTT
30	-2842	TTCAAAACTA	GGTGCCCAAA	TTCTTGGATT	TGACCACATA	AGAGAGCTTT
	-2792	ATCAAGAAGA	TCAAGAACTC	TCATCCATCT	ATGCCCAATG	TCTACATAGA
	-2742	GCACAAGGAG	GTTACTATGT	GTCCGAGGGA	TATCTTTTTA	AAGAAGGAAA
	-2692	ACTTTGCATT	CCCCAAGGAA	CACATAGAAA	ACTCCTTGTC	AAAGAATCAC
	-2642	ATGAAGGGGG	ACTCATGGGG	CATTTTGGAG	TTGATAAAAC	TCTAGACTTT
	-2592	TAAAAGCAAA	ATTTTGTGGG	CCACACATGA	GGAAAGATGT	CCACGACATT
35	-2542	GTCTAGAGTA	TCTCATGTTT	AAAAGCAAA	TCTAGAACAA	TGCCGCTGGA
	-2492	CTCTACACCC	CTTTGCCGAT	TGCAAAGCTC	CTTGTGAAGA	CATTAGCATG
	-2442	GATTTTCATTT	TAGGACTTCC	TAGGACTGCA	AGAGGCCATG	ACTCTATCTT
	-2392	TGTGGTAGTG	GACCGTTTTA	GCAAAATGTC	TCACTTTTAT	CCATGCCACA
40	-2342	AAGTAGATGA	TGCTCAAAAT	ATTTCTAAAC	TCTTCTTTAG	AGAAGTGGTG
	-2292	AGACTCCATG	GTCTCCCTAG	AAGTATAGTG	TCCGATAGAG	ATCACCTTAA
	-2242	ATATATAATT	ATACACTTGT	TTTTTTTCTC	TTTTTTATTT	TATCAAGTAA
	-2192	AAAGTATTTG	TTCTAGATTA	TTATGAGTAT	ATACTTACTT	TCTGTATTTT
	-2142	ATTTCTTTCT	ATTTTTTATG	ACGATGAAAT	TTCTTATTAT	ATCCAGACTT
45	-2092	TTCATATATA	TTTTTATTTT	TTTTCCATCT	AGATGCTCTG	TACTTTTCTT
	-2042	CAGTTGAAAT	TTCCACTCTC	CAACAAAACA	TCATTCAAGT	TTTGTATAAC
	-1992	ACTGTGACGT	TAACCAGTTA	AAATAAGAAA	ATCATGTAAT	ATAAATTATT
	-1942	TCAGTAGATA	TTTTAGAATT	ACAAATACGA	TAAATAATTA	AATTTAAAAA
	-1892	ATTATTAAC	AATGAATTTT	TTTGAAATTT	AATATAAAAC	TTAGACTTGT
50	-1842	GGTTTCTTCA	TTCAAGTCAA	ACCTTTTCTT	ATTGTGTGGC	GTGTGCGTGA
	-1792	ACATCGAATT	TGGGTGCTTT	ATGCCGCTTT	ATCTTCATCT	GCACCTTCAA
	-1742	ATTAATAATT	TAATCCGGA	AAATAATAAA	CCCACACACT	GTTTTATGCA
	-1692	TATATTAAGA	TAAATAAAAG	AGAACTATTT	TAAAGAATAT	AAAATAATAA
	-1642	ATGTAACAAA	TGATGTCACT	AAAGAAGAAA	AAAATTAAAC	AGAATTGTAA
55	-1592	TATATTTCTT	TATGAAATGT	TTTGTGCATT	ACCGAGAGAG	GTGCAACATG
	-1542	ATACACGCAA	GCATCTAACT	AGTTTGGTAA	TTCTTTTCA	ACATCGnTAA
	-1492	GCACATCACA	CTAAAATTAC	TTTAAATAGA	TAAATTAGAT	TCAATTGGAT
	-1442	GACATTAATT	TATAATACTC	TATCCAAAAT	TATAACTATA	AATAAAAAGT
	-1392	TATTTTTAGA	AAATAAGTAA	TGAAAATTTA	ATTCTAAAAT	TTATAACACT
60	-1342	TTTATGCTGT	GTTTGTTCG	AAGCATAGAA	AAATAAAAAG	TTATTGTTGG

5
 10
 15
 20
 25
 30
 35
 40
 45
 50

```

-1292 GAATGAAAAG TGAAGAAAAT CATGTAATAA AAACAAAATG ACACGACAAT
-1242 CAAAAAAGAA GTTTTCATGC AAAACTTTTT TCAAAATTTA CACTTTTATG
-1192 ATGTGTTTGT TTCGAAGTGT AGAAAAACGA AAAGTTATTA TTGGTAATGA
-1142 AAAGCGAAGA AAATCACGTA ATAAAAACAA AGCAAGATGG CACGACAATC
-1092 AAAAAAAGT TTCTACACAA AACTTTATTC AAAATTTTACA ACACTTTATAT
-1042 GTTGTTGTTT GTTTCCGAGG TATAGAAAAA CAAAGAATTA GTGTTGGTAA
-992 TGAAAAGTGA AGAAAACCAT GTAATGAAAA CAAAATGGCA CGACAATCAA
-942 AAAAAGTTTT CACGCAAAAT TTTCTTCAAA ATTTATAACA TTTTCATGTT
-892 GTGTTTGTGT CAAAGCCTAG AAAAACGAAG AGTTACTATT GGTAATGAAA
-842 AGCGAAGAAA ACCACATAAT AAAAACAAAA TGGCACGACA ATCAAGAAAA
-792 AGTTTTTCACA CAAAACTTTT TTCAAAATTT ACTATGTTTA TTTTCGAAATT
-742 TAGAAAAACG AAGAGTTATT ATTAGTAATG AAAAGCGAAG AAAACTACGT
-692 AATAAAAAAC AAAATGGCAC GACAATAAAA AAAGTTTTCA CGCAAAATTT
-642 TCTTGGTGCG CAGAAAGTTA TATATATTAA TTAATTAATT TTCATTTACT
-592 TTTTCCCTT TTTATTTTAA AGTTAAATTA TTATTATTTT CATTTAAAAAT
-542 ATAAATATTA TTTAAATATA AAAAATATAA CCTTAATCAA AACAAAGCCT
-492 TAATCTAAAA TTTACAACAC TTTTAACCTT AAAATTAACT TTAAAAGGAA
-442 AATGATAGTG TGACAACTAA AAAAGTTGTA TACAACCCTG TCATAGGTTT
-392 AGAAATAAAT ATATATAATA AAGAGTAAAT TTGTAATTAA ATGATATAAA
-342 AAAGTATTAA AATAATAATA TTTAGAGTAG TAATATGGTT GTATAAAAAA
-292 ATGTGGTTGT CCATATATCA TTATTCACCT TAAAATATCA TGACAAATAT
-242 TTTCACCGAA AGATGGAAAG AACGAAAAGA GCGTTGGATA ATGGAAAAAT
-192 ACAAGCAATC TCCCTCCAGT ACTTTGCATA ACATTTTGTG TTAGTGATGA
-142 GTTTTTTATC ATATATATTT AGAATATAGG AAAATTTTAG AATCACGTGG
-92 ATAGCTATAT AATAGTAATA TTTTAATTTA TAATGTAGTT GATTTTATTT
-42 GTCAACTGGT ATACATAAAT ATGTGTTGAT AGTGGGTGAC TTGTGGCTTA
9 AAGAAATGTC CAGAGGCTGA CAACAACCTCT GCACAGACTA GCGTAAAC
57 ATG AAG TCC AAT TTT GCT ATT TTC GTA GTC TTT TCT CTT CTT CTT
1 M K S N F A I F V V F S L L L
102 CTG GTACCTCTTCAATCTTCTCTACAAAACTCTGTTGCTCTTTCACCTCTGTTGTA
16 L
160 ATTTTGTTTACACTTTTGGAAAAATTGAAGCTGATATATATGTAACAACCTTTTCAGTTTT
219 GTCTGCACTGAAACTGATAGAAAAATATACGTTTTGTGGATATATATAG GTT GGC
17 V G
274 AGT TGC AGC TGC GCA AGA AAA GAC ATG AGA GGG TAT TGG AAG GAT
19 S C S C A R K D M R G Y W K D
319 ATG ATG AAG GAG CAA CCT ATG CCA GAA GCA ATC AAA GAC CTT ATT
34 M M K E Q P M P E A I K D L I
GAG GAT TCA GAA GAA GTG TCA GAA GCA GGG AAG GGT CGT TTT GTT
49 E D S E E V S E A G K G R F V
AGG GAC TTC GAT GTA AAG CCT AAT GTC ATA TTA TAT CAC ACA CAT
64 R D F D V K P N V I L Y H T H
GTT GTG CCC ATG AAG CAG AGG CAG AAG AAT AAA GAT TGA
79 V V P M K Q R Q K N K D •
493 AGACTATGTGATTGGCAGTTTCAGACTTATTTGGCACCAAATTTATGATGCTCTTGTGCTG
555 TTTCAAATTTGTACTCAAACCTTTGAACCCCTTGCAGCATCTTGCTTCTTTTTGGTCTTGCT
617 GAATTTTGTACAGTTATACTGTACGAATAGTTTCTCTTCATAATAAGCAACCTTTCCTCT
679 C
  
```

FIGURE 3

5 101001 CAAAACAAAAGCAAATGCCGGTTTTCTTATTATTATTTCGAACCTTTAGAC
100151 CTTTTTGTAACGTTTCTTTAATTTTTTTCTTGATAAAGAACCCTATTAT
100201 ATCTTAGCTAAATATTTACCTCATTTTGTATTATGAGCTAAACCACCCCAA
100251 AAATATTGTAGTTTTGCTTTCCGATTAACTGCCAAGCAAGTGATTAGAT
100301 ATATTAAAGGAAAAATGAATGAAAGGACAAAAAATATAAACGACAATATT
100351 TGAATACTGATATTATCTCCATTCTCAAATATTTTTGATTTATTGTGAC
10 100401 AATATTTGGTTGTTTCCCATTTGCTACATCTTTGAGGACATGAAATGATA
100451 ACATATATATGAACGAGTATAATACATTCTCGTTTCATTTTACAAATAAT
100501 GTCATTTTATGCTAACATTTTTTAAATTTAAAAATTATCCTTATAAGATTTC
100551 AGTGTATTATTTTTACCATGGTACTGTAAAGTCGGATGCTATATATATATA
100601 TATATATATATATATCAAAAATGACACTGAAGAATTTATTTGAACATAAAA
15 100651 CTAAAAACGTAAAAATAAAGAATTTTCAAAAATCAAAAATTTTATATA
100701 AAAATATAGATAAAAATGTTAATATAGTACAACCTTCTATTCAAACAGAGAG
100751 AATAAATCTTCTATAGACAGTGAATATCCATTATAATAACGAGCAATAGT
100801 TGAATGTTGCAGTACAAAAGAGAATTGTAATATTTGTGCATGATTGAG
100851 AAATCTAAGTTGACTTTGAATTTAAAGGCTAATTCCAACAAGTACATGTA
20 100901 GAAGTTGACTATAGCTATATATTTTACTACAAATTGATCATTTCAAGAAAG
100951 ACATTTAAATTAAGATATGCATGCATGACTTGATTGAACCCCACTCGCTT
101001 GCTTCGTGCCATTGACACAAGATGTTACTTTTAAATGCAAGGTAAATTATG
101051 GATATACTCTTCTGTATTTTTTGTAGTAGATATTTTTACGAAAATTGTTT
101101 TTTTTCAAAATCAAATGATATTTTAAATTTTCAATATAGAATTAATTA
25 101151 AATTTTAATTAATTTTGAAGATTTATATGCTGCAGATTAGATTACCATTG
101201 GTGAAATCATGTTTAGGTAAATAATAATGATGTTGTAGTTTAGGAAAAA
101251 AAAAAATTCCTTAACTCTTTATGTAAGAATGTTAACTTCAATTATAAAAA
101301 TATGAAGCAGTATTATATAAGATGTTTAACTAATCGAATAATATTTTTTG
101351 GGATGAAATTTTCTGTCATATGTTTCTAAAAAATAATATGTGAAAAATT
30 101401 AACATTCATTGTATGTTTATAAGAAATATATGTGAGTTTGTGTTAGATAA
101451 ATAATACTTAAATTAAGAATTTGTAAGATTATCTGCACCTTCAAATATG
101501 TTATTTTTTCTTTTATTTAAAAATATCAGCAACATTCTAAATGATTTTAT
101551 TTTCTTTAAAAAATTGAAAAAATGAAATTAGCAAATATGTAAATTTTAA
35 101601 ACGAATTTAAGAAAAAACTTTGTAAAGATATGATATGCTTTATAAAAAA
101651 ACTTGGTGGCGTACCTACTAAATATGATCACATTAGAGATTGTATCCTT
101701 TAGCATATAGTATGTAGTATAGATATCTATATTTTTATTATTAAAGAGC
101751 ATATTCATAATATAGGTATTATATGTTAATTACAATAAACGTTCAATTG
101801 TTATGTTAGTTTTTAGAAAACTTATTGCGTGTGCATATCAATGTGAGAAA
40 101851 GCGACTCCACATGTGAGATGTTGGTCTGAGAAAGCTTCTGCACTTGGTC
101901 GGAACACTTTCATGGACTAGAATGCAATCCATCTATTCAAAGAAAAGCAG
101951 TTGTCCATGCATGCCTCGGTTTTTTCACATTTGGAAGCAGCGCAACAATGT
102001 CTTACATAATATGCGATCGATCACTCTGCAACCAATATTCAAGTACATAG
102051 ACCATGACATCAAAAACATTATCACACCGAGAAGAAAGAAACGTCAATTT
45 102101 GGTAACCTAATGGCGTTATGCCTGCGGTGAATTCTCCTAAGAGTTCTCCC
102151 AAAATTTTATTGATTCCTTGTTTTTAACTTTTTTCGCCAAAGAATCATACAT
102201 ATAGATTTGACACCATTTCACCTTATCAAATACAAGTGAATAAATAATTT
102251 CAAGCTTGAAAGGAATTTAATCATGATCTAAACCTAAACGACAAATCTT
102301 CACAAGTGAGAATCACTAATTGACTACCCCTTGGTTCGCATATACATCATT
102351 GTTGTAATCTGAAAATTGGTTTGGATTGATCTGATATGTCATTCATAT
50 102401 AAAACTTGTATTATTTATTTTAGAATTTGCCGCAACAGATAAATCATC
102451 ATCTATTTAGAAAAATTTTCATTTGCACCACAATTAATCAGGGGAAAAGGT
102501 GAAATCACATATCTTATCTACACTCTTATTAAATTAACGCCATAATATA
102551 ACAAATTTTCAAATACCACTTATGAGAAGCACTAAGATCACCTTTTTCTT
102601 TATGACTTTCTTTCTAAAGCTAAGCTGGTAGTCATGACTCATGATTATCC
55 102651 TTTTCTAATGGGAATATTGTGGAAGCGGTTTCAAATCTTTAGACAAAAAT
102701 TCCATGGCCACTAAAAGTTAGCAAAGTTAAATAAGTTTAAAAAATATG
102751 AGTGTAAGTTGGCCATATGCCATATTGTTGAGATCATAACAAGAGAAATAA
102801 TAGTTTATTGAAGTTTAGATCATAATACAATACATCATTGCCTTCATCA
102851 ACATTTTCCATGGATTGAGAGGATCAACTTCAATACTAATGGTGGGGTC

102901 TTATTCATCCATTGCTCTCTAGCCAATTAAGCAGTTAGGTTATTTGTGTA
102951 CTCTAGTAGTTGCCAAATCAATCTTAATATTCACAATGTGTAATTTCTA
103001 ATTACGTATAGATAAATGACTAGATAACACGTGGCTTTGGTTTTATCAGG
103051 AAAGTTTTCCAAATCATATATATGAATGTAGAATAGTGTCTTCATTAAT
5 103101 TATTAATTAGCATCTCACCATCTGAGACTGGGAGCATGTGACAAGTTGAC
103151 ATGTGTATTAAGAGAACTTTGAGAAAACCACTTTTATGATACTCCCATCT
103201 GAGACTGGGATGAGTACCATTTTATAAAAAATATGAGTAGTGAAAAAATAT
103251 TCAAAAAAATTCCTAACATGTCCTTTAAAACATTTTAACTTATAATTTT
103301 AACAAACATCTTCCAATATGCGTTATGAAAACTTTATAAACTTTTTTAT
10 103351 AACATGCTTTTGGAAATTTTATAAATCTGTATTTTGGAAAACAAAGTGAT
103401 ACTTTTGGAAATAGACAAATGAAGTGCTATTTTAAAATTGATATCATA
103451 AGTCTTAACTGTGGTTTGTGTTGAATTTATTTATATACTGTCAAAATAA
103501 AACTAAATAAATAAATAAATTATTTTATAATCATGAAGATAATATTATC
103551 ATAAAAAGATAAATAAATAAATCAACAAATTTATATTTGTTAATAAAAAATAC
15 103601 TTTGAGCTCTTCTTCATAAGACTTTTCCAGCTTCCATCTAGAAAATCACA
103651 TAAATTAAGATAAATAAACCAGTAATAACATAGTTCACATTTCTAATCTT
103701 AGTCTTAGATTTGTTTAAATTTTCAAAGGTTTAGGTATTGTATATGTTT
103751 TTTTATTGGGTTGCTAGATTTTGATCCAAGAAGAAATGACGGGTTGTAGT
103801 ATAGATGGTTTGTGTTGAGTTTTTCCCCTTGGTTTACTTCGTTTGGTTTT
20 103851 TGTCCCAGAAATGTTCTTGTACTCGCTGGTTTATGTCTCTACAAAGTCC
103901 ACCACCATTGCCGGCTCTTTGTATTTCAACTTGAATTCTAAATTCGATTG
103951 ATGAAAAAATGTATCTCTTAAAGTCCATTAGTACCAAAAAATACTAT
104001 ATCATTACTACATAAAATAGTCTTGGGTTTTCCAAAGTATTTGTTGATA
25 104051 TATGTTAAGAGTTTCAAAATAGACACATAGATATAATGTTGAAATGGGACC
104101 TCTCACATAATTATCTCCTTTTCTCTTCATTTCTCTACCTCTCAAGTTTC
104151 CAATCCCACCCTAAGGTAATTTATTTCTTAACCTAAGTAAATTTGTAAAC
104201 AAATCTTAACTAGCTACAAATGTGTATTACAAGTCTTAAATAAAAACCTA
104251 CTTTAATTCAAAGGTATTAAACCTTCTTAAATTGATACTTACTTAGTATC
30 104301 GATCGGTCTAGTTTAGGGTTTGGACAACACACCATCATGGGGACGAAATT
104351 AGTCATTCTACGGTGTCCAAGACACAAATCTCGGACTCGATGTGGATATG
104401 AACCTTCATTATAACTTTTAACTTCATAAAAACTAACTATTAGGAGGAAG
104451 AATCGGAATCTGCATATCAATCACATAGACTATAGTATACTTAGATTTT
104501 GATCTAATCAATGGGCTCCTTCAACTAATAAGTAGCCCACTACCAATAAT
104551 GAAATCATAAGACATTATTAAATTAATCAATGTTCTAAAAATACTTTGGT
35 104601 TATGTGTCCCGTAGAGCTAATGTGCACACACAATGAAAGTTGACCCGTTT
104651 CACTGTGCCACTTTTATGATCTTTTCTTTTAGGTTAAATCCAACCTTTTA
104701 TAATCTCATCTTGTATCAAACAAAACCTTTGGCCTGTCTTTTTCATAAT
104751 TTAAGTAACCTCTACGGAGAAAAGCCAACATTTTCTTCTGTTTTATTCT
40 104801 TTTTAAAGAAAAATGAATTCAAGGGGACCCCAAATTTAAAGGAAAACCA
104851 AAACCTCCTTTCTATGTATTTATTACTTGAAGTTTTCTATGTAATCAACAA
104901 TCCTAACAGTAGAGAATAAAAAACATCGTTTTGGGAGGTTTTATATTAGC
104951 ATATGAGAATAGTTCTAAAATTGTTTACACAAAAATTAGATTTTCTTTT
105001 CCTCTGTCAATGGAGCTATATCACTTGTCATTTTGCTTAACCTTTGCGG
105051 GAAGATTGTTATGAAACAGTTTTAATGGAATTCAGTTGCCAATGTCACG
45 105101 TTTAATATGTTTTGTCCCTATACTTTATTGAATCTTATAATCTTTGTTAT
105151 AGAATTATCTACTTTTAGTATTTTACATTAAACATAATCTATAGAATTCTT
105201 CTTTGTCTATACAATTAACAAGTAATATATTCTTAATACATATTAAAA
105251 ATGGTGGTGTGCTATCTGAGCTGTAATAGTTGATTGCTCCAGAGAAGAA
50 105301 TAGACAAAAATCCTTACTTAAGAGGCCCACTCTGAAAAATTTAGACAA
105351 GAAAAATTAAACAAAATTAGGTTACACATATTATCATTTATATATATGCA
105401 CAACACAAAGTTGACCTTGCAATGTACTATTGAATAAAAAATAAATGTC
105451 AAGAAGAGAGGGAATTATCACTGTTACCAAGAAAACAACTTCTCTAAAC
105501 AGGTCTCTATATATATAAACTTTAACACCTAAAGAATTAACACAGATCAA
105551 GAAAAAATCCTCAAAAACAAAGTTAAAGCAGAC ATG AAG CAA CAG CAA
55 1 M K Q Q Q
105599 CGT TAC TTG GTC GTC TTC ATC GTC CTT TTA AGC TTT CTT
6 R Y L V V F I V L L S F L
105638 CTG GTAAAGCTTCTTCCTTAATTATATTAAACCCTAATTAAGATCTCATATA
19 L
60 105691 TCTGAATGTTGTATATATTTGTTGGTATAG TTT GTG AAT CTG AGT
20 F V N L S

FIGURE 4

1 AAGCTTTACAAATGTCCCCAAAGATGAAACCACGTTATTATTAGTAAATCCTGAAAAGG
61 TTAACGCTTCTGTTCTCCTCGAATTCCTAAACCATCTGAAATATCTAGTGGTTTAAATGGAG
5 121 ACTTGAGGATATAGTCTCCTGAACCAGCTGTACGGCTGAGTTAGATAACATTACTGAAT
181 TTCTACGGGAGCGGTTGAAATCACATTCGCCCCCTTAAGAAGAAGCCTACACCGGGCACC
241 TTCTTTACGCAATTCGAAATTTAGTCTTGCCAGGCAGTCGTTGGATCGAAGGTCTTTTTTC
301 GATACCGAGGAATCTGACTTTGCAAGGAATAATTCCTAATCACACCACCCCAACCCCTGA
361 ATACACTTCAGGACCCCTCTGAAACCAACTTCGTTTCGGCTAAATCACAAGAATCTCCCAC
10 421 TCATTCCGATTTTAGCCAATTAAATATGATATCGGTCTGGGAAGCCGATAAGGAAATTCT
481 ACAAAAAGAGTTTATGAATGAGGAAAATAAGGAAAAGAGAGAACTATTTTTTAGGTACCC
541 TGAAAGAGAACGAGAAAAATTTAGAAAAAATACTACTCTCATCTGTACTGTTCACAAA
601 GAATATCnnnnnAATGGTTAGATATAAAGAAAAGGATAAGTATGATTAAACTGAAAC
661 CACGTGCGCAGAAACAAAGTGAATTCCTCCCTTTAGAGGAAGTTCGTTTCTTAAATATAG
15 721 AAAACAAAGAAGTAGTCGCCTCCCTTTTAAATGATCTCAGAAAAACGAGAAGTAAGTA
781 TAAAAGATATTCAAATCTACACAGTCAACTAAATTTTACTAATCAAATGCTTTTTTCAAT
841 TAGCAAATAAAAAACAAAAGAAAAAGmGAAAATTGAAGAAAAATCGTTAATAAAACCAT
901 TTAAATTCTCAGAAGAAGAGATAAAAAAGTTAAAAATTGGTCAAACCTTTGGATTCTTTAT
961 ACGATGAAGTAAACAAAAGTTATCTATCTCGGTAATAAAAGAAAAACCGAAATCTAATA
20 1021 ATGATATGCCCAAAGGACAAATCCAAATCAAGAAAGTTTATAGACGAAATCGAAAAGAT
1081 TAAAACAAACCTCTGAACGACACAATAAATGTGATAGAAGAACTAAAACTCAGACTCAT
1141 GTTCAGAGTCTCCCGATCGTATTGAAAAAATAAAACGTAATAAATCAGAGATTTCCAGTA
1201 AGCCGAAATTTTACTACTCGCCCCACCTTCGATATCATCGAGATGGCGATGGACACCTCA
1261 GCATTGATGGAATGATACTGAGTGATATGATGGATGACAGATGATGAATATAGAAAAAC
25 1321 TCACGAAATAACAATGGCCGCTACAGCATATAGAGTAAAAACATACCGAGGAACAAACAAT
1381 AAAATTAATTATATCTGGATTACGGGAGTATTAAGAGGCTGGTGGGATAATTACCTCAT
1441 GCCAGAACAAAAGAAATTATGTTCTAAGCTGTGTAAAAATAGAAAACGAGAAGTAATAC
1501 ACTAATGGTGGAACATTTGGTGGTAGCAATAATTCATAACTTTATAGGAGATCCAAAGAT
1561 TTTTGAAGAAAGAACATCTTTATTACTTCATAATCTAAGATGTCCAACCTTAGGTGACTT
30 1621 TAGATGGTATTGAGAAAATTTTTTAGCTATGGTTTTAACAAGGGAAGATTGTAGAGAACC
1681 TTTCTGGAAGAACGGTTTATAGCTGGATTACCGGATATCTTTGCTGAAAAGGTAAAAGA
1741 AAATTTACAAAAGGAATGCCCAAACACCCAATTAAAAGATGTACCATACGGGAAAATAAG
1801 TTCAGTTGTAAAAATACAGGTCTTCAGTTATGCAATAATATGAAAATAGAAAATAAGAT
1861 AAAAAAGAGTGAGAGTCAGGGCATCAAGGAATTAGGGGAATTTTGTACTCAATAGGTTA
35 1921 TGAACGAAATACCCCTCCATCAAAAAATAAAAAAGAAAAATAGCAAAAAGAAACAGGGAG
1981 AAACAAGCGCTAAAACAAGCGCTAAACCAGCACGTAAAAATTTTAGAAAAACGGTTAATT
2041 TTAGAAAACCATGAAAGTCTAATGATAAGCCCACTATAGTCTGTTATAAATGTGGACGCA
2101 TAGGACACATGAAGCGAGACTGTAGACTAAAAGAAAAAATTAGTAATTGACCATAAGTG
2161 ATGAATTAAAAGAACAAATGGA AAAA ACTTCTGATAAATTCCTCCAGAAGAGGAAGAAACA
40 2221 GAAGAATCAATAGGAGATTCTGATTACGAAGTATTGGACATGAGGATAACAATTGTAATT
2281 GTGTCTATAAAATAAATACGATAAGTAGTGAATTAAAATTTGCGTTAGATTGCTGATA
2341 AAATTAATAATCCGAGGAAAAGACCAAGCCTTAATAGACATGAAAAGGCTACTCGTTG
2401 AAAAAAGATGAACCCAGTTCATCTTCAAAAAACCTGAATTTATAGGATATGATTTTAAAG
2461 AAATATTGAGAAAAGCGAAAACATCACATAAAGAAATAACCATTAGCGATCTTAATAGTG
45 2521 AAATAAATAAATTAAAAGCCGAAATCGAATCTATAAAAGTCGAGCTACAAGAATTAAAAG
2581 ATAAAATTATACATGAGGAATCCATCTCCTCTGCCGACGAAAATTCACAAGAAGAGGAAG
2641 CTAGTAGACCTTCCATCAAAGAAATAACATACAAAAGACAAAAGTGGCATGTA AAAATAG
2701 CCCTAGAATTTGTTTGTGTTTGTGACCGTTTCATTGTGGTCAAAGATGAGTCCTTACCTAA
2761 CACAATAAAAAACGTTACTCTTAAATATCAAAGGAGAGCTACAAATATCAATGAATGAAT
50 2821 GACATTAATATTTTTCTTTAGTTTTAAAACCTGAATGAGTTGTTTTATATAATATCTGAC
2881 TGACTGACATTTTTATTTTTCTGAAAATGAGGAAGGTTTATTACGTTAACACCATATAT
2941 ATATTTTTATCTCAAAGTCAACGAAATATTATAAAAGAATCAATTAAAAAAAATTATTCT
3001 TTTGCAGAAAAAAAATTA AAAATATGAAACTCCTCCACACCATATTACCATATTATAAA
3061 TATAAAAAACCTCTCACAAATGTGCATTCTGGAATCTTTATGTTGAGAGATTAATCTC
55 3121 TAAAGAAAAAAGGTTGAGAAAGGTGCAGCAACA **ATG** TCT CCA TTC TGT AGA
1 M S P F C R
3172 AAC TTT TCA ATG GCA TGG GTG CTT ATG GCA TTT GTG TTG TTT
7 N F S M A W V L M A F V L F
3214 GCA AAC AGT GCT ATG CCC ACA AAT GGA TCC ACT GTT GGG GTA

21 A N S A M P T N G S T V G V
 3256 AAA AAC ATG TTG GGT GGT AAA TTG ATG CTA AAC GTT TTA TGT
 35 K N M L G G K L M L N V L C
 3298 CCC CAT ATT GAT AAG CAA CAC ATT ATC CCG AAT GGT GGT TCA
 5 49 P H I D K Q H I I P N G G S
 3340 TTT GAG TGG AAG TAC AAT GGT GGT GCT CCA CCA ATA GGA CAA
 63 F E W K Y N G G A P P I G Q
 3382 TCA CCA TTC ATG TGT TTC TTT CGG TGG AAT AAT GTT CAT CAC
 77 S P F M C F F R W N N V H H
 10 3424 TCC CTT GAT CTG TGT TCA CCA AGC AAG TAT ACT GGT TGT GAA
 91 S L D L C S P S K Y T G C E
 3466 AAT GCC ATT TGG GAA ATC AAA GAA AAG CAA TTT TGT AGG TAC
 105 N A I W E I K E K Q F C R Y
 3508 AGA GGT GGA CCT ATT AAT TAT TTT TGC TAT GAC TGG GAT GAT
 15 119 R G G P I N Y F C Y D W D D
 3550 TAG TTATATAGATTATTCATGTTTCATCTCAATAAAAAAATGACTTTAGAGTGATTCTT
 3609 AGTTTGCTTAACATTCTTACATATTCCTAACTATTCCGTCACCTACCACCCGTAACCTATAT
 3669 TTATTTTAAATTAGTATCTGTACAGTTTTTATTTTAAAAAAGGTTATGTGGATTAGAAG
 3729 AGAGATAAATATGTAGACGGTCACCAACCTTAATTTTGAACCTATGTAAGACTATATTGA
 20 3789 CCAAGAATATATGTTTAAACTCATTCATTTAAAGACTATATCTCCATTTATGATTATGCA
 3849 AATGCAATTAGTTTTTTTTTTCATTGAAGAATCAAAGAAAGTTATCATTAAAAAGTAT
 3909 CATTAAATCACTTATATGTTGTTTCTTAATATCCTTATTGTTAATAGAATAATTTTTTTT
 3969 ATCCTTTAATTAAGGTTATTACTACTTTTTTTTCATATCTTCATTATTTTGAAATATTTT
 4029 TAAAATTTTATCAATTTTTTGTAACACCCAGAAAATACATGTAACCTATCACTTTTTTTT
 25 4089 TATTACAAATTTATGACTTATAGAAATACAAATATTAATAATAAGGTTCAAACTACA
 4149 TCCTAAAGTCTTTCAGACCCTCTGACACATGTATCATCTGCTCGTATATGTGATACAGTC
 4209 ATCGCAGTTCACAAGATAACAAGAAAACCAAGGTAAGCTAATGAAAAAAATTCCTATA
 4269 CATATTTAATTCATGCAAAAAGAACAGTCAAAGTAATCATTTATAAACATTTCTTTAAA
 4329 TATTGTTATATAAAATTTCAATATCAATTTTCATCATTCATATAGACCACACATGGATCTA
 4389 TTTTCAATCACAATCATTGGATTTCATTTTAATCCTACTTCGnCTTCCAGAAGACTCATT
 4449 AAGTATGCCCCCTACCAGAGACTAACACCTAATCAAAGAGAAATGATCAAGGTAAGTTCAA
 4509 ACATCCAATAACGAGTGCCCTACAGTGGGACCCAATGTGTATGAACTCCTTATCAGCTTCT
 4569 CACCACCTGATATCTTATTCTATATGACGTAGATCATCAGTGAACTAGAGGATCTCCGT
 4629 TAAACATATGTTTTTTTATACCTAATGTGCATCAAACAACAACTCACACATTATCCCAAATG
 35 4689 TATGACATCAATTTTCATACAATTTTCATCATTCATATATAATACATATCATTGAATCACA
 4749 TAACATTTAAAAATTCATACCATTCAAGAACTTTTCCAACATCAAAGCAATATTTACTT
 4809 TCAAACCTATCAAATATAATTATTATTTAATAAAGCTt

FIGURE 5

	142000	TTATCTTATTTCCATATAAATGTTGTTTTACTTTCAAAATTTTAAATTTT
	141950	TTATATTTATCTTTTTACAGTTTAAAATTAATAAAATGAACTTTTTTTC
5	141900	TTAAATGTGTAAAAATATAAAATCAAAAAAGTTGTATATGGTACATGGC
	141850	ACAATCTTATAAAATTATTAATTTGAAAACGATACTTTATATAATAAAATT
	141800	ATCTTAGTTGACATTTTTATTAGTGTTTTCAATCATATTTTTGTTTGCTT
	141750	GATAAGCGTAAAACAAATCAAACCTAACGATACTTTATATAATAAAATTA
	141700	TCTTAGTTGACATTTTTATTAGTGCTTCAATCATATCTTTGTTTGCTTG
10	141650	ATAAGCGTAAAACAAATCAAGTAAAGTTGGGCACCTCAATTGTTTTAAAA
	141600	AAGTTTGGGTACCTCAAAAAATTAATAGGTCTTGTCAGATTCTTACAAAA
	141550	AAATCTGGAAGAATTTATGAAAGAAGGGGGGGGAGGGGGGAGGGGGGGG
	141500	AAGTGAAGATGAATATTCAACAAAAGAGGGTAGGCATGATGTTAAGTGAG
	141450	TTAAAAAATATGTTAATGGAGACAATTTCTGTAAACAAACCCGTTAAT
	141400	TGAAAACGATAGCATTCTTCTTAACAATGTAAAACGATATTGTTTATC
15	141350	ATAACTACTCATTAAATTTCTGAGTTTCAAATCATATAAAGATTTAGGGG
	141300	GGTGATTCAATTAAGGATTTGAAATGATTTGTATTAAATGACAAATCC
	141250	CATGTTATTTCAAACATGAATTGTAAAACTTTTTTAAATCAAGTGTTA
	141200	TTAGATTAGTGATTTTAAAAATGTACAACCAAACCACTGTTATTGGAAAC
	141150	ATTTTAAGTAGTGATTTTAAATGACTTGAGTGATTTTGGGTGGGATTGC
20	141100	AGAAAATTTCTTAGTTAAGAAATCAACATCCAAATCTCATGGTTTCAAG
	141050	TAGAATTTGGGAGAATTTTAAATAACAAATCTCCTAATTTACCAAAAGTCA
25	141000	CCAAAATCATTTAAAAACTCATTTAAATTTAAATGATTTCAAATCTCCAG
	140950	TTGAATACATCCCTTGGAATTAGAGATTTTGCTCGATTTGGGACCTAAG
	140900	ATTGAATTTTGGGGATTTAGTTTAAATCGTTACAACAAAATGACATCGTAT
	140850	TATTGTTATAGGAAACAATGTCTTTTCAGTTGACATGTATGTTAATAGA
	140800	AAATTAACCTTATTAACGGGATTTGCTAACCCATTTAACATCGTAACTAA
	140750	ATGGTCAAGTCAATAAAAGTTTGGTATTTATTTGAAAAGTCAACGTAAAGT
	140700	TTGATATTTATTTGAAAAGTCAACATAAATTTGATATCTTATTTTCGTTTC
30	140650	GACAGACATAAGGATTTACATCAATGTTTTTAAATAAATTAAAGATTATTA
	140600	TGACATTTTTCATTAAATTTGCCAATGTTTTTCGAAACCAAGATACTC
	140550	AAAATTGACATACCTAATTCAATCTACATTTGTTTGACAGCAATTCACGT
	140500	GCCTTGACCACATGGCACATACTGGCAATACATCAATTTTAAGGAAAAGG
	140450	TAGATTCGGATACAATATAATGGAATAAGTGGAAAGGATCATTGACTAC
35	140400	TTGACTTGTAACAAACAACACACAGTATATAACTCATTCGACATTTACAA
	140350	ACAACATTGTGCTAGCTTAAACTCCCTCTCCTATTCAAAAAA ATG
	1	M
	140305	GAT ATT CCA AAG CAA TAT CTA TCA CTA TTC ATA TTG
	2	D I P K Q Y L S L F I L
40	140269	ATT ATC TTC ATA ACT ACA AAA TTA TCA CAA GCC GAC
	14	I I F I T T K L S Q A D
	140233	CAT AAA AAC GAC ATT CCA GTT CCC AAC GAT CCA TCA
	26	H K N D I P V P N D P S
	140197	TCA ACA AAT TCT GTG TTT CCT ACC TCG AAA AGA ACC
	38	S T N S V F P T S K R T
45	140161	GTG GAA ATC AAT AAT GAT CTC GGT AAT CAG CTA ACG
	50	V E I N N D L G N Q L T
	140125	TTA CTG TAT CAT TGT AAA TCA AAA GAC GAT GAT TTA
	62	L L Y H C K S K D D D L
	140089	GGT AAC CGG ACT CTG CAA CCA GGT GAG TCG TGG TCT
50	74	G N R T L Q P G E S W S
	140053	TTT AGT TTC GGG CGT CAA TTC TTT GGA AGG ACG TTG
	86	F S F G R Q F F G R T L
	140017	TAT TTT TGT AGT TTT AGT TGG CCA AAT GAA TCG CAT
	98	Y F C S F S W P N E S H
55	139981	TCG TTC GAT ATA TAT AAA GAC CAT CGA GAT AGC GGC
	110	S F D I Y K D H R D S G
	139945	GGT GAT AAC AAG TGC GAG AGC GAC AGG TGT GTG TGG
	122	G D N K C E S D R C V W
	139909	AAG ATA AGA AGA AAC GGA CCT TGT AGG TTT AAC GAT
60	134	K I R R N G P C R F N D

	139873	GAA	ACG	AAG	CAG	TTT	GAT	CTT	TGT	TAT	CCT	TGG	AAT	
	146	E	T	K	Q	F	D	L	C	Y	P	W	N	
	139837	AAA	TCT	TTG	TAT	TGA	CAACAATATGCTGATGTTCTGCTCTTTTAC							
	158	K	S	L	Y	•								
5	139793	GACTCATGGAGTTTTCATTGTTTGTGAACAATAATATAAAACATATAAAATT												
	139743	TCTATTATTCCAAGTTCCAACTTATAATAAATTTGATAATCATATCATATT												
	139693	ATCATCTTAAGCATTCAATGCTACAAAGATAATACCCCCAAGCTATTTTA												
	139643	CATTA AAAAGCTGAAACAGAGACACAATACTAACGATAAAAGTTTCGTAGTA												
	139593	TCTTTATGCAACCATACATACATATACACAAAGATAGACAGGTAGTGTCC												
10	139543	TAATAATCTTACTTGGGTGAGGTATGAACAGCAGCAGCAGCAGTAGATACCAT												
	139493	TGTATTCATACACACATATTATGAGGCCCTGCGCAAGTTTTGTAGTAAC												
	139443	CATGCTCTCCCCACATCGCTCCCCACGAGTTCTTGATAATCCAA												

FIGURE 6

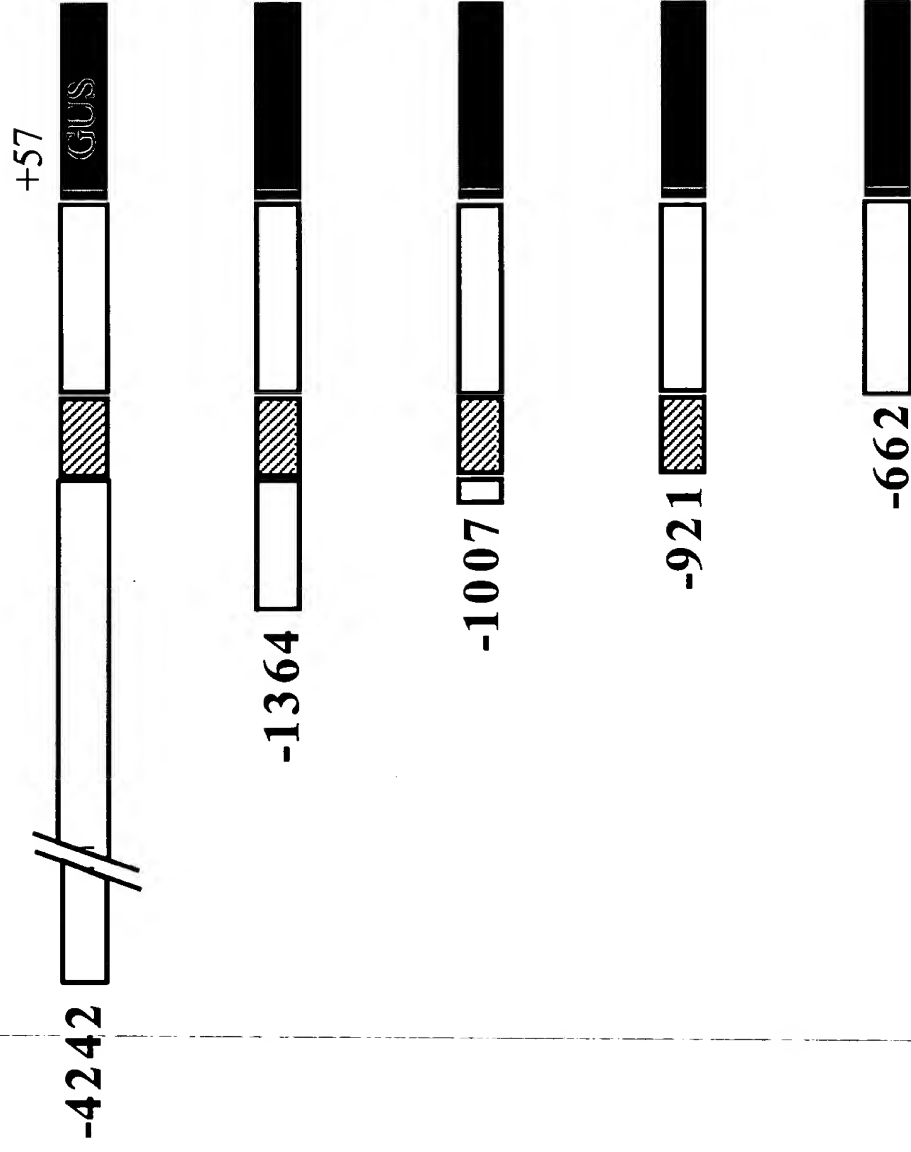


Figure 7

G564 promoter: Gain of function constructs

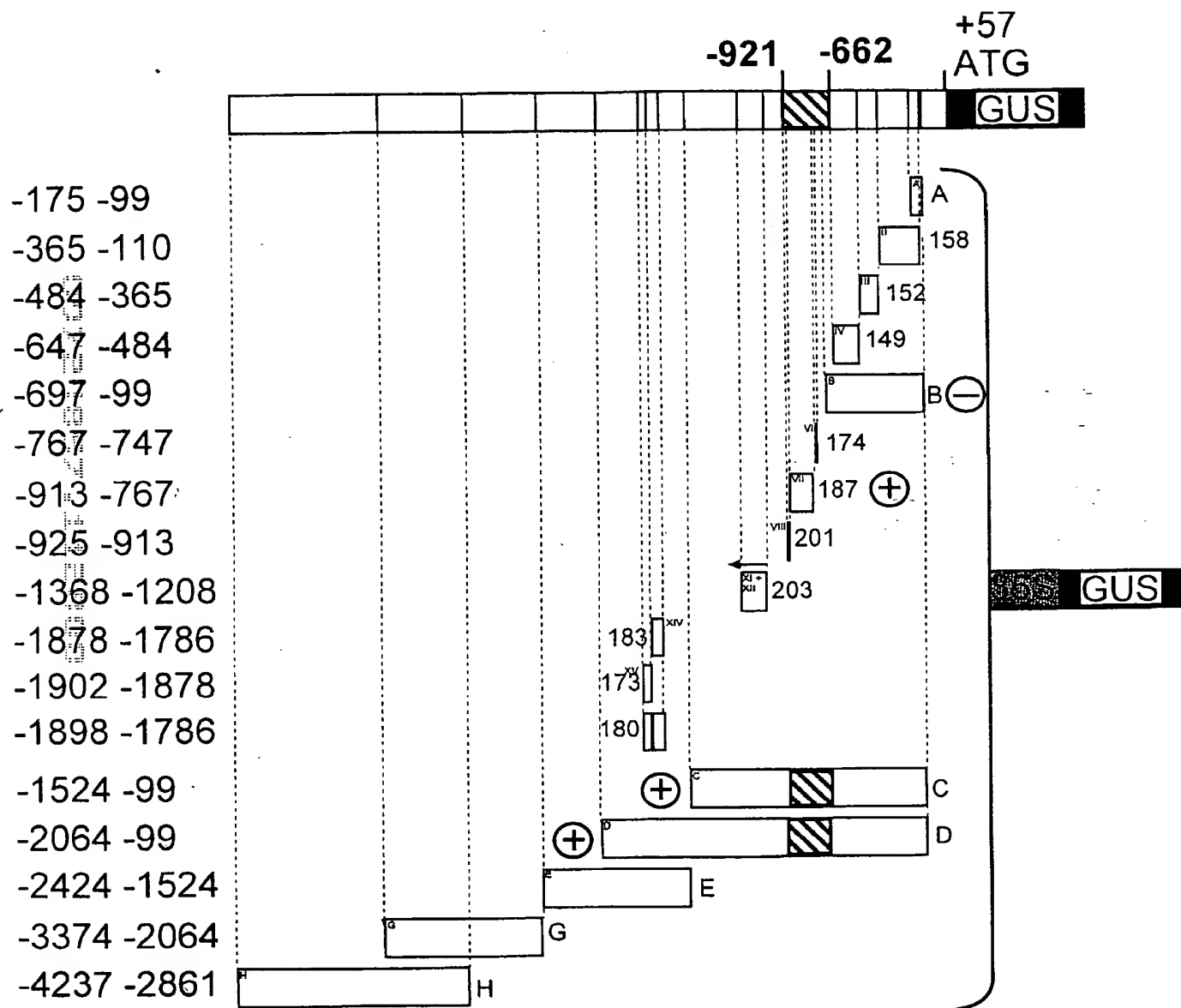


Figure 8

Web Signal Scan Program

Database searched: PLACE

URL: <http://www.dna.affrc.go.jp/hdocs/PLACE/>

This is the sequence you submitted

>G564 promoter (-921 to -662), 450 bases, 3D1A0BF4 checksum.

```
TGAAAAGTGAAGAAAACCATGTAATGAAAACAAAATGGCAGCACAATCAA
AAAAAGTTTTCACGCAAAATTTTCTCAAAATTTATAACATTTTCATGTT
GTGTTTGTTCAAAGCCTAGAAAAACGAAGAGTTACTATTGGTAATGAAA
AGCGAAGAAAACCATATAATAAAAACAAAATGGCAGCACAATCAAGAAAA
AGTTTTCACACAAAACTTTTCAAAATTTACTATGTTTATTTTCGAAATT
TAGAAAAACGAAGAGTTATTATTAGTAATGAAAAGCGAAGAAAACACGT
AATAAAAACAAAATGGCAGCACAATAAAAAAGTTTTCACGCAAAATTT
TCTTGGTGCGCAGAAAGTTATATATATTAATTAATTAATTTTCATTTACT
TTTTTCCCTTTTATTTTTAAAGTTAAATTATTATTATTTTCATTTAAAT
```

Notation: H = A, C, or T

R = A or G

K = G or T

W = A or T

RESULTS OF YOUR SIGNAL SCAN SEARCH REQUEST

/tmp/signalseqdone.9437: 450 base pairs

Signal Database File:

Factor or Site Name	Loc.(Str.)	Signal Sequence	SITE #
-300ELEMENT	site 1 (+)	TGHAAARK	S000122
2SSEEDPROTBANAP	site 101 (-)	CAAACAC	S000143
ACGTABOX	site 296 (+)	TACGTA	S000130
ACGTABOX	site 296 (-)	TACGTA	S000130
AP3SV40	site 159 (-)	TGTGGWWW	S000169
CAATBOX1	site 44 (+)	CAAT	S000028
CAATBOX1	site 189 (+)	CAAT	S000028
CAATBOX1	site 323 (+)	CAAT	S000028
CAATBOX1	site 138 (-)	CAAT	S000028
CANBNNAPA	site 101 (-)	CNAACAC	S000148
CCAATBOX1	site 138 (-)	CCAAT	S000030
CEREBLUBOX2PSLE	site 55 (-)	TGAAAAC	S000033
CEREBLUBOX2PSLE	site 201 (-)	TGAAAAC	S000033
CEREBLUBOX2PSLE	site 333 (-)	TGAAAAC	S000033
DOFCOREZM	site 4 (+)	AAAG	S000265
DOFCOREZM	site 53 (+)	AAAG	S000265
DOFCOREZM	site 112 (+)	AAAG	S000265
DOFCOREZM	site 149 (+)	AAAG	S000265
DOFCOREZM	site 199 (+)	AAAG	S000265
DOFCOREZM	site 282 (+)	AAAG	S000265
DOFCOREZM	site 331 (+)	AAAG	S000265
DOFCOREZM	site 364 (+)	AAAG	S000265
DOFCOREZM	site 419 (+)	AAAG	S000265
DOFCOREZM	site 216 (-)	AAAG	S000265
DOFCOREZM	site 399 (-)	AAAG	S000265
DOFCOREZM	site 408 (-)	AAAG	S000265
GT1CONSENSUS	site 120 (+)	GRWAAW	S000198
GT1CONSENSUS	site 141 (+)	GRWAAW	S000198
GT1CONSENSUS	site 196 (+)	GRWAAW	S000198
GT1CONSENSUS	site 253 (+)	GRWAAW	S000198
GT1CONSENSUS	site 69 (-)	GRWAAW	S000198
GT1CONSENSUS	site 90 (-)	GRWAAW	S000198
GT1CONSENSUS	site 347 (-)	GRWAAW	S000198
GT1CONSENSUS	site 388 (-)	GRWAAW	S000198
GT1CONSENSUS	site 436 (-)	GRWAAW	S000198

GT1CONSENSUS	site	218 (-)	GRWAAW	S000198
GT1CONSENSUS	site	401 (-)	GRWAAW	S000198
GT1CONSENSUS	site	402 (-)	GRWAAW	S000198
MAMMALENHAN	site	158 (-)	GTGGTTTK	S000121
MARTBOX	site	324 (-)	TTWTWTTWTT	S000067
MRE1	site	356 (-)	TGCRCNC	S000068
NTBBF1ARROLB	site	418 (-)	ACTTTA	S000273
POLASIG1	site	168 (+)	AATAAA	S000080
POLASIG1	site	301 (+)	AATAAA	S000080
POLASIG1	site	324 (+)	AATAAA	S000080
POLASIG1	site	237 (-)	AATAAA	S000080
POLASIG1	site	411 (-)	AATAAA	S000080
POLASIG3	site	268 (-)	AATAAT	S000088
POLASIG3	site	427 (-)	AATAAT	S000088
POLASIG3	site	430 (-)	AATAAT	S000088
POLASIG3	site	433 (-)	AATAAT	S000088
POLLEN1LELAT52	site	11 (+)	AGAAA	S000245
POLLEN1LELAT52	site	119 (+)	AGAAA	S000245
POLLEN1LELAT52	site	156 (+)	AGAAA	S000245
POLLEN1LELAT52	site	195 (+)	AGAAA	S000245
POLLEN1LELAT52	site	252 (+)	AGAAA	S000245
POLLEN1LELAT52	site	289 (+)	AGAAA	S000245
POLLEN1LELAT52	site	362 (+)	AGAAA	S000245
POLLEN1LELAT52	site	71 (-)	AGAAA	S000245
POLLEN1LELAT52	site	349 (-)	AGAAA	S000245
PYRIMIDINEBOXHV	site	400 (+)	TTTTTTCC	S000298
RAV1AAT	site	97 (-)	CAACA	S000314
ROOTMOTIFTAPOX1	site	374 (+)	ATATT	S000098
SEF4MOTIFGM7S	site	170 (-)	RTTTTTR	S000103
SP8BFIBSP8BIB	site	134 (+)	TACTATT	S000184
TATABOX2	site	81 (-)	TATAAAT	S000109
TATABOX3	site	375 (+)	TATTAAT	S000110
TATABOX4	site	368 (-)	TATATAA	S000111
TATABOX5	site	238 (+)	TTATTT	S000203
TATABOX5	site	412 (+)	TTATTT	S000203
TATABOX5	site	434 (+)	TTATTT	S000203

For more information about the SignalScan Program, please contact Dr Dan S. Prestridge Tele: (612) 625-3744 Advanced Biosciences Computing Center, E-mail: danp@biosci.umn.edu 1479 Gortner Ave. University of Minnesota St. Paul, MN 55108 The TFD data is at the gopher site, gopher://genome-gopher.stanford.edu. For more information about the WebSignalScan service, please contact Meena Sakharkar, meena@biomed.nus.sg, BioInformatics centre, NUS .

Database Searched: PlantCARE
URL: <http://sphinx.rug.ac.be:8080/PlantCARE/>

Sequence submitted:

>G564 promoter (-921 to -662) 11/21/00

+ GAAAAGTGAA GAAAACCATG TAATGAAAAC AAAATGGCAC GACAATCAAA AAAAGTTTTC ACGCAAAATT
+ TTCTTCAAAA TTTATAACAT TTTCATGTTG TGTGTTGTTT AAAGCCTAGA AAAACGAAGA GTTACTATTG
+ GTAATGAAAA GCGAAGAAAA CCACATAATA AAAACAAAAT GGCACGACAA TCAAGAAAAA GTTTTCACAC
+ AAAACTTTTT TCAAAATTTA CTATGTTTAT TTCGAAATTT AGAAAAACGA AGAGTTATTA TTAGTAATGA
+ AAAGCGAAGA AACTACGTA ATAAAAACA AAATGGCACG ACAATAAAAA AAGTTTTCAC GCAAAATTTT
+ CTTGGTGCGC AGAAAGTTAT ATATATTAAT TAATTAATTT TCATTACTT TTTCCCTTT TTATTTTAAA
+ GTTAAATTAT TATTATTTTC ATTTAAAA

- CTTTTCACCT CTTTGGTAC ATTACTTTTG TTTTACCGTG CTGTTAGTTT TTTTCAAAAG TCGGTTTTAA
- AAGAAGTTTT AAATATTGTA AAAGTACAAC ACAACAAAAG TTTCGGATCT TTTTGCTTCT CAATGATAAC
- CATTACTTTT CGCTTCTTTT GGTGTATTAT TTTTGTTTTT CCGTGCTGTT AGTCTTTTTT CAAAAGTGTC
- TTTTGAAAAA AGTTTTAAAT GATACAAATA AAGCTTTAAA TCTTTTGTCT TCTCAATAAT AATCATTACT
- TTTGCTTCT TTTGATGCAT TATTTTGTG TTTACCGTGC TGTATTTTTT TTCAAAAGTG CGTTTTAAAA
- GAACCACGCG TCTTTCAATA TATATAATTA ATTAATTTAA AGTAAATGAA AAAAGGGAAA AATAAAATTT
- CAATTTAATA ATAATAAAAG TAAATTTT

3-AF1_binding_sit

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
3-AF1_binding_sit	ST	260	+	1.000	0.860	AAGAgttatt

Function:

AAGAA-motif

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
AAGAA-motif	Avena sativa	6	+	1.000	0.903	gtgAAGAA
AAGAA-motif	Avena sativa	151	+	1.000	0.870	gcgAAGAA
AAGAA-motif	Avena sativa	284	+	1.000	0.870	gcgAAGAA

Function:

ABRE

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
ABRE	Hordeum vulgare	293	+	1.000	0.854	actACGTaat

Function: cis-acting element involved in the abscisic acid responsiveness

ACE

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
ACE	Petroselinum crispum	293	+	1.000	0.908	actACGTAat

Function: cis-acting element involved in light responsiveness

.AE-box

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
AE-box	Arabidopsis thaliana	67	-	1.000	0.852	AGAAaatt
AE-box	Arabidopsis thaliana	345	-	1.000	0.852	AGAAaatt
AE-box	Arabidopsis thaliana	361	+	1.000	0.852	AGAAagtt

Function: part of a module for light response

AT1-motif

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
AT1-motif	Solanum tuberosum	409	+	1.000	0.859	ttttATTTtaaa

Function: part of a light responsive module

Box_4

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
Box_4	PC	375	+	1.000	1.000	ATTAat
Box_4	PC	379	+	1.000	1.000	ATTAat
Box_4	PC	383	-	1.000	1.000	ATTAat

Function:

Box_I

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
Box_I	PS	107	+	1.000	1.000	TTTCaaa
Box_I	PS	203	+	1.000	0.857	TTTCaca
Box_I	PS	219	+	1.000	1.000	TTTCaaa
Box_I	PS	240	+	1.000	0.857	TTTCgaa
Box_I	PS	241	-	1.000	0.857	TTTCgaa
Box_I	PS	249	-	1.000	0.857	TTTCtaa

Function:

Box_II

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
-----------	----------	----------	--------	-------------	--------------	----------

Box_II	ST	139	+	1.000	0.889	TGGTaatga
Box_II	AT	161	+	1.000	0.954	CCACataat

Function:

CAAT-box

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
CAAT-box	Hordeum vulgare	43	+	1.000	1.000	CAAT
CAAT-box	Arabidopsis thaliana	137	-	1.000	1.000	aCCAAT
CAAT-box	Hordeum vulgare	188	+	1.000	1.000	CAAT
CAAT-box	Hordeum vulgare	322	+	1.000	1.000	CAAT
CAAT-box	Arabidopsis thaliana	351	-	1.000	0.857	aCCAAG

Function: common cis-acting element in promoter and enhancer regions

ERE

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
ERE	Dianthus caryophyllus	239	+	1.000	0.875	ATTTcgaa
ERE	Dianthus caryophyllus	241	-	1.000	0.875	ATTTcgaa
ERE	Dianthus caryophyllus	413	+	1.000	0.875	ATTTtaaa
ERE	Dianthus caryophyllus	441	+	1.000	0.875	ATTTaaaa
ERE	Dianthus caryophyllus	442	-	1.000	0.875	ATTTtaaa

Function: ethylene-responsive element

G-box

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
G-box	Zea mays	17	+	0.842	0.870	CATGta
G-box	Zea mays	38	+	1.000	0.903	CACGac
G-box	Zea mays	94	+	0.842	0.886	CATGtt
G-box	Zea mays	183	+	1.000	0.903	CACGac
G-box	Zea mays	317	+	1.000	0.903	CACGac

Function: cis-acting regulatory element involved in light responsiveness

GC-repeat

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
GC-repeat	Oryza sativa	351	-	1.000	1.000	gCACCaag

Function: ?

HSE

Site Name	Organism	Position	Strand	Core simil.	Matrix simil	sequence
HSE	Brassica oleracea	49	+	0.944	0.878	aAAAAagtt
HSE	Brassica oleracea	50	+	0.944	0.912	aAAAAgttt
HSE	Brassica oleracea	52	-	0.944	0.874	gAAAActtt
HSE	Brassica oleracea	66	-	1.000	0.978	aGAAAattt
HSE	Brassica oleracea	77	-	0.833	0.868	aTAAAtttt
HSE	Brassica oleracea	87	-	1.000	0.853	tGAAAatgt
HSE	Brassica oleracea	196	+	0.944	0.912	aAAAAgttt
HSE	Brassica oleracea	198	-	0.944	0.874	gAAAActtt
HSE	Brassica oleracea	210	+	0.944	0.874	cAAAActtt
HSE	Brassica oleracea	212	-	0.944	0.912	aAAAAgttt
HSE	Brassica oleracea	213	-	0.944	0.878	aAAAAagtt
HSE	Brassica oleracea	327	+	0.944	0.878	aAAAAagtt
HSE	Brassica oleracea	328	+	0.944	0.912	aAAAAgttt
HSE	Brassica oleracea	330	-	0.944	0.874	gAAAActtt
HSE	Brassica oleracea	344	-	1.000	0.978	aGAAAattt
HSE	Brassica oleracea	361	+	1.000	0.888	aGAAAgtta
HSE	Brassica oleracea	385	-	1.000	0.853	tGAAAatta

Function: cis-acting element involved in heat stress responsiveness

I-box

Site Name	Organism	Position	Strand	Core simil.	Matrix simil.	sequence
I-box	Pisum sativum	93	-	0.857	0.883	aACATga
I-box	Pisum sativum	162	+	0.857	0.883	cACATaa
I-box	Solanum tuberosum	163	-	1.000	1.000	tATTatgt
I-box	Pisum sativum	237	-	0.857	0.941	gAAATaa
I-box	Pisum sativum	367	-	1.000	1.000	tATATaa
I-box	Pisum sativum	372	+	1.000	0.941	tATATta
I-box	Pisum sativum	391	-	0.857	0.941	tAAATga
I-box	Pisum sativum	411	-	0.857	0.883	aAAATaa
I-box	Pisum sativum	423	+	0.857	0.883	tAAATta
I-box	Solanum tuberosum	424	-	1.000	0.903	aATAattt
I-box	Arabidopsis thaliana	426	-	1.000	0.863	aATAataat
I-box	Arabidopsis thaliana	429	-	1.000	0.863	aATAataat
I-box	Solanum tuberosum	431	+	1.000	0.951	tATTattt
I-box	Pisum sativum	433	-	0.857	0.883	aAAATaa
I-box	Pisum sativum	439	-	0.857	0.941	tAAATga

Function: part of a light responsive element

P-box

Site Name	Organism	Position	Strand	Core simil.	Matrix simil.	sequence
P-box	Oryza sativa	406	+	1.000	0.857	CCTTttt

Function: gibberellin-responsive element

Prolamin_box

Site Name	Organism	Position	Strand	Core simil.	Matrix simil.	sequence
Prolamin-box	Oryza sativa	145	+	1.000	0.913	tgaAAAGc
Prolamin-box	Oryza sativa	278	+	1.000	0.913	tgaAAAGc

Function: cis-acting regulatory element associated with GCN4

TATA-box

Site Name	Organism	Position	Strand	Core simil.	Matrix simil.	sequence
TATA-box	Daucus carota	79	-	1.000	1.000	TATAaatt
TATA-box	Brassica juncea	80	-	1.000	1.000	TATAaat
TATA-box	Helianthus annuus	81	-	1.000	1.000	TATAaa
TATA-box	Brassica oleracea	82	+	1.000	0.908	tTATAac
TATA-box	Brassica napus	83	-	1.000	0.892	gtTATA
TATA-box	Oryza sativa	117	+	0.818	0.912	TAGAaaa
TATA-box	Oryza sativa	169	+	0.818	0.872	TAAAAaac
TATA-box	Zea mays	248	+	0.909	0.879	TTTAgaaa
TATA-box	Oryza sativa	250	+	0.818	0.912	TAGAaaa
TATA-box	Oryza sativa	302	+	0.818	0.912	TAAAAaa
TATA-box	Oryza sativa	325	+	0.818	0.912	TAAAAaa
TATA-box	Daucus carota	364	-	1.000	0.863	TATAactt
TATA-box	Brassica juncea	365	-	1.000	0.857	TATAact
TATA-box	Zea mays	366	-	1.000	0.879	TATAtaac

Site Name	Organism	Position	Strand	Core	simil.	Matrix	sequence
TATA-box	Oryza sativa	367	-	1.000	0.956	TATATAa	
TATA-box	Oryza sativa	368	+	1.000	0.929	TATAtat	
TATA-box	Oryza sativa	369	-	1.000	0.929	TATAtat	
TATA-box	Solanum tuberosum	370	-	1.000	1.000	TATATA	
TATA-box	Glycine max	372	+	1.000	0.891	TATAtt	
TATA-box	Oryza sativa	407	-	0.818	0.872	TAAAAaag	
TATA-box	Zea mays	413	-	0.909	0.879	TTTAAaat	
TATA-box	Zea mays	442	+	0.909	0.879	TTTAAaat	

Function: core promoter element around -30 of transcription start

•TC-rich_repeats

Site Name	Organism	Position	Strand	Core	simil.	Matrix	sequence
TC-rich_repeats	NT	7	-	1.000	0.952	gTTTTcttca	
TC-rich_repeats	NT	68	+	1.000	1.000	aTTTTcttca	
TC-rich_repeats	NT	152	-	1.000	0.909	gTTTTcttcg	
TC-rich_repeats	NT	191	-	1.000	0.885	tTTTTcttga	
TC-rich_repeats	NT	248	-	1.000	0.914	tTTTTctaaa	
TC-rich_repeats	NT	285	-	1.000	0.909	gTTTTcttcg	
TC-rich_repeats	NT	346	+	1.000	0.915	atTTTcttgg	

Function:

WUN-motif

Site Name	Organism	Position	Strand	Core	simil.	Matrix	sequence
WUN-motif	Brassica oleracea	18	-	1.000	0.948	tCATTacat	
WUN-motif	Brassica oleracea	139	-	1.000	1.000	tCATTacca	
WUN-motif	Brassica oleracea	237	+	0.857	0.948	tTATTtcga	
WUN-motif	Brassica oleracea	242	-	1.000	1.000	aAATTtcga	
WUN-motif	Brassica oleracea	272	-	1.000	0.948	tCATTacta	
WUN-motif	Brassica oleracea	296	-	0.857	0.948	tTATTacgt	

Function: wound-responsive element